CLIPS at TREC-11: Experiments in Filtering

Christophe Brouard
Laboratoire CLIPS
Equipe MRIM
BP. 53
38041 Grenoble cedex 9, France
Christophe.Brouard@imag.fr

Abstract

At the TREC9 conference, we presented a new adaptive filtering system called RELIEFS. This system which is based on the idea of resonance, combines for each term t, the relative frequency of relevance knowing t and the relative frequency of t kwowing relevance. On the basis of other experiments, several changes have been made. We improved our threshold adaption, we slightly changed our relevance evaluation function and we gave up the use of conjunctions and thesaurus. The system is now focusing more exclusively on the combination of both reverse frequencies that we believe to represent the fundamental aspects of relevance estimation. This year we used the system in its new version and we tested it on the Reuters corpus. Focusing on the combination of the two frequencies, we varied their relative importance. The results show globally a good effectiveness especially when both frequencies are balanced.

1. Introduction

In a logical approach, it has been found that the evaluation of relevance of a document D to a query Q can be based on the evaluation of the implication from document to query D->Q (van Rijsbergen, 1986). Some authors also emphasized the role of the reverse implication (Q->D) (Nie, 1988). These two entailment relations can also be found in most probabilistic models, formulated as P(Q/D) (Probability of Q knowing D) and P(D/Q) (see Crestani, Lalmas, van Rijsbergen & Campbell, 1998, for an overview of Probabilistic models).

If one considers a document as a set of terms, and a query as a specification of what we are looking for, the entailment D->Q may be decomposed to the judgment of "if the term is present then the document is relevant" for each term of the document. So, if we use a set of documents with the relevant judgments associated, the entailment D->Q can be decomposed in a set of relative frequencies of relevance knowing term for each term of the document. Similarly, the reverse entailment relation can be decomposed in a set of relative frequencies of term knowing relevance. From a pragmatic point of view, the use of the first frequency is easily understood since it allows to favour the terms which predict relevance (when the term is present then the document in which it occurs is relevant). The consideration of the reverse frequencies may be justified by the fact that it allows to favour terms which have been met many times in relevant documents. Since they appear relatively frequently, these terms are likely to be usefull for the next evaluations, and we can think that their presence in relevant documents is not a coincidence (comparing to rare terms which occured once or twice).

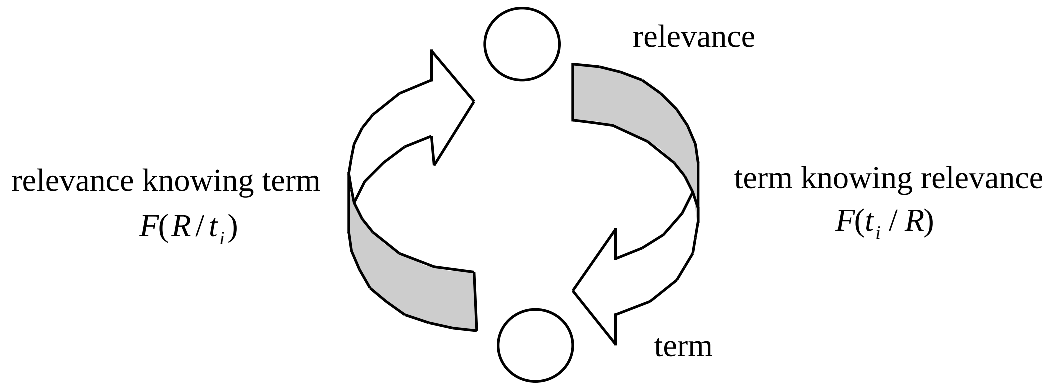

Figure 1 – Relationships between terms and relevance

We believe that both entailment relations capture the essence of relevance. As we noted before they are found in the most IR models but not always in a obvious way. With the system RELIEFS we try to isolate them and to use them in a simple way. So, for each term t_i, we propose to compute the frequency of relevance knowing term $F(R/t_i)$ and the reverse frequency $F(t_i/R)$ (Figure 1).

2. System description

The RELIEFS document processing can be decomposed in three steps:

1. Selection of N terms from the document,
2. Estimation of the document's relevance,
3. Revision of term's relative frequencies.

2.1 Step 1 : Selection of N document terms

All the document terms are compared with the terms which have been extracted from the query, from the document examples given for learning and from the documents which have been previously selected. They are sorted by the value the product $F(R/t_i) * F(t_i/R)$. If less than N terms can be selected in this way (in our experiments we choose N=30), this selection is completed by the document terms in their lecture order.

2.2 Step 2 : Relevance estimation

Considering the terms which appear in the document, the score of the document is computed as follows :

$$\frac{\sum_{i=1}^{N} F(R/t_i) * F(t_i/R)}{\sum_{j=1}^{N} F(R/t_j) * F(t_j/R)}$$

where i are the indices of the N best terms (according to the product $F(R/t)*F(t/R)$)) which are present in the document and where j are the indices of the N best terms (present or not present in the document). In RELIEFS, the relevance of a document can be interpreted as a kind of resonance in a network (Figure 2) in which each relative frequency corresponds to a weighted connection and each term corresponds to a node (Brouard & Nie, 2003). A relevance node is built for each query.

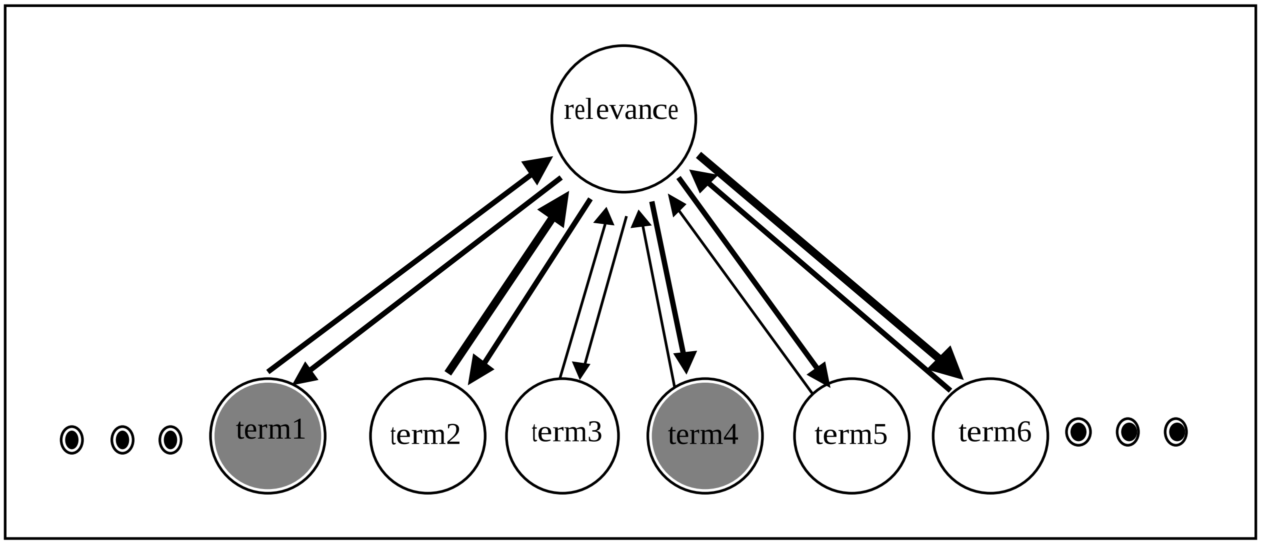

Figure 2 - Relevance evaluation in RELIEFS of a document in which term1 and term4 occur

2.3 Step 3 : Relative frequencies updating

If the evaluation of step2 is larger than a defined threshold the document is selected and the relevance feedback takes place. Given the relevant judgment, the N terms selected in the first step are submitted to an updating process on their relative frequencies.

2.4 Threshold adaption

The queries are very different from each other. Therefore, a different threshold is determined for each query. The value of the threshold is determined empirically. The initial threshold is computed on the basis of the score of the three first relevant documents. We used the following strategy to adjust the value:

- If a selected document is irrelevant, the system is considered to be too tolerant. The threshold is increased.

- If a document (that we do not know if it is relevant or not) is not selected, the system is considered to be too strict. Then the threshold is decreased

The increase scale is set to be higher than the decrease scale because there are much more unselected documents than the selected ones. The decrease of threshold in the second case is due to the fact that we do not want the system to remain silent for a too long period. This allows to gradually correct an initial threshold that is fixed too high. In both cases, we considered different criteria for the modification of the threshold, including:

- The number of irrelevant documents that are selected consecutively: The higher this number, the larger the increase scale and the smaller the decrease scale.
- The number of consecutive relevant document: The higher this number, the larger the decrease scale. This change only concerns the decrease case.
- The number of documents considered: The larger this number, the lower the change scales. The intuition behind this criterion is that we would make larger changes at the beginning of the filtering. When a certain number of documents have been treated, the system should stabilize.

We also considered a more global criterion on the probability of relevance. Each score is associated with a frequency of relevance and we try to adjust the threshold on the score for which the probability is 0.33 (this selection criterion is optimal for the utility measure).

3. Experiments

3.1 Preprocessing

We used the Porter's stemmer to get the root form of every term and we removed the stopwords.

3.2 Details of processing

All parameters (which mainly concern the threshold tuning) were tuned on the Oshumed corpus (TREC9). For each document of the Reuters corpus, we only considered the title and the text fields. For updating frequencies, the unjudged documents were considered as irrelevant. A pseudo-relevance feedback has been applied when the scores were smaller than a defined threshold. In this case, the document is considered as irrelevant.

3.3 Varying the relative importance of the entailment relations

In our previous tests, both entailment relations are considered with the same importance: they are multiplied. Our question is, should they play equal role in the estimation or should one factor be more important than another? In order to answer this question, we vary the importance of one of the implication by replacing the weight $F(R/t_i)$ by $F(R/t_i)^{\rho}$. The factor ρ changes the relative importance of this relation. The relevance estimate is defined as follows:

$$\frac{\sum_{i=1}^{N} F(R/t_i) * F(t_i/R)^{\rho}}{\sum_{j=1}^{N} F(R/t_j) * F(t_j/R)^{\rho}}$$

3.4 Results

Globally, it turns out that the best results are obtained when the two frequencies are balanced (ρ in [1.0,1.4]) even if we can observe that good results for ρ=1.6 and ρ=1.7 (Table 1). This is consistent with the previous observations we obtained with the Oshumed corpus.

ρ	Utility	ρ	Utility	ρ	Utility	ρ	Utility
0.0	-0.26	0.5	10.35	1.0	16.68	1.5	14.01
0.1	2.65	0.6	12.25	1.1	17.25	1.6	16.82
0.2	4.92	0.7	12.42	1.2	17.84	1.7	17.27
0.3	8.01	0.8	13.22	1.3	17.68	1.8	13.19
0.4	9.41	0.9	15.44	1.4	16.95	1.9	12.03

Table 1 : The impact of ρ on the utility.

We submitted the best run (ρ=1.2) for comparison on utility criteria. The comparison is favourable since about 80% of the scores are above or equal to the median. Like the other systems we obtained better results for the first 50 queries than for the 50 last ones which are built differently.

In conclusion, results suggest that the two reverses frequencies are sufficient clues to estimate the relevance. Furthermore, the fact that the best results are obtained when the two aspects are balanced should indicate that they are both necessary and equally important.

References

Brouard, C., & Nie, J.-Y. (2000, 11/00). The system RELIEFS: a new approach for information filtering. Text Retrieval Conference (TREC-9), Washington D.C.

Brouard, C., & Nie, J.Y. (2003). Relevance as Resonance : A new Theoretical Perspective and a Practical Utilisation in Information Filtering. *Information Processing & Management*. Revision under review.

Crestani, F., Lalmas, M., C.J, V. R., & Campbell, I. (1998). Is This Document Relevant?...Probably: A Survey of Probabilistic Models in Information Retrieval. *ACM Computing Surveys*, 30(4), 528-552.

Nie, J. Y. (1988). An outline of a general model for information retrieval. Proceedings of the 11th Annual ACM Conference on Research and Development in Information Retrieval, Grenoble.

van Rijsbergen, C. J. (1986). A non-classical logic for information retrieval. *The Computer Journal*, 29(6), 481-485.